Journal of Voice
Vol. 14, No. 1, pp. 22–33

Comparison of Acoustic and Perceptual Measures of Voice in Male-to-Female Transsexuals Perceived as Female Versus Those Perceived as Male

*Marylou Pausewang Gelfer
†Kevin J. Schofield

**University of Wisconsin—Milwaukee, Wisconsin, USA*
†Milwaukee Public Schools, Milwaukee, Wisconsin, USA

Summary: The present study explored significant differences between male-to-female transgendered speakers perceived as male and those perceived as female in terms of speaking fundamental frequency (SFF) and its variability, vowel formants for /ɑ/ and /i/, and intonation measures. Fifteen individuals who identified themselves as male-to-female transsexuals served as speaker subjects, in addition to 6 biological female control subjects and 3 biological male control subjects. Each subject was recorded reading the Rainbow Passage and producing the isolated vowels /ɑ/ and /i/. Twenty undergraduate psychology students served as listeners. Results indicated that subjects perceived as female had a higher mean SFF and higher upper limit of SFF than subjects perceived as male. A significant correlation between upper limit of SFF and ratings of femininity was achieved. **Key Words:** Transsexual voice—Gender identification—Acoustic analysis.

Among adults with communication disorders, a small but fascinating population includes transsexuals, or transgendered individuals. A transgendered individual strongly believes that his or her true psychological gender identity is not congruent with his or her biological or physical gender. Many of these individuals live for years trying to conform to the social role specified by their biological gender, but eventually seek medical and surgical help, as well as other forms of counseling and therapy, to achieve the physical characteristics and social role of the gender they feel to be their true one.

The transgendered individual seeking sexual reassignment goes through an intense and often difficult process to become socially accepted as a member of his or her psychological sex. In addition to achieving a physical appearance and bodily configuration consistent with the new gender, the attainment of appropriate speech and voice characteristics is also a desired goal for the transgendered client. Unfortunately, the voice of the male-to-female transsexual (one whose biological gender is male but who wishes to become female) remains in the male pitch range, despite treatment with feminizing hormones that bring about the development of other female characteris-

Accepted for publication December 1, 1998.
Address correspondence and reprint requests to Marylou Pausewang Gelfer, PhD, Department of Communication Sciences and Disorders, University of Wisconsin—Milwaukee, P.O. Box 413, Milwaukee, WI 53201, U.S.A.

tics. To develop a voice and also a communication style appropriate to the reassigned gender, a speech-language pathologist is often consulted.

Which vocal characteristics are most important in gender identification, and how successfully can the perceived gender of a speaker be changed? Bralley, Bull, Gore, and Edgerton[1] reported marginally successful results following seven hour-long therapy sessions with a 49-year-old male-to-female transsexual who had undergone both hormone treatment and sex reassignment surgery. The subject had a speaking fundamental frequency (SFF) of 145 Hz prior to therapy, and 165 Hz following therapy. In addition to measures of SFF, the investigators also conducted a perceptual study. Fifteen listeners rated the masculinity or femininity of conversational speech samples from 3 male control subjects, 3 female control subjects, and pre- and post-therapy samples from the transgendered subject. Results revealed that, on a 7-point masculinity-femininity rating scale, with 1 representing a very masculine voice and 7 representing a very feminine voice, the client improved from a rating of 3.7 pretherapy to a rating of 4.6 post-therapy. Male and female control subjects were rated 2.0 and 5.9, respectively. The investigators concluded that even after therapy the voice of the subject was still discernible from that of biological female speakers. However, the investigators did not report whether listeners identified the subject's voice as belonging to a female either before or after therapy.

Mount and Salmon[2] also attempted to raise a post-reassignment surgery transgendered client's speaking fundamental frequency. However, they also included data on increasing formant frequencies as well as SFF, specifically F_2 in vowels, by altering tongue placement. Therapy was conducted in 88 hour-long sessions over an 11-month period. At the end of treatment, the subject had increased her SFF from 110 Hz to 210 Hz in prolonged vowels. Measures of F_2 values for /i/, /ɑ/, and /u/ also revealed an increase in frequency. Mount and Salmon reported that the subject achieved an optimal SFF for a female speaker after 4 months of therapy but, according to subject report, was still perceived as a male over the phone. After 6 months, when the F_2 values began to increase, the subject reported that she was finally being perceived as a female over the telephone. This perceptual information, however, was not part of the study, but reported to the authors by the client. Clearly, additional perceptual evaluation by listeners unaware of the nature of the speaker would have helped to determine the importance of formant frequency cues to gender identification.

Spencer[3] extended the work of previous researchers by including a total of 8 transgendered speakers in her study. Spencer's subjects were all receiving hormone treatment and were living either part- (50% or more) or full-time as females. Not all had completed their reassignment surgery; however, all were perceptually attempting to use "female" speech patterns, in the judgment of the investigator. In this study, listener judgments of both the gender of the speakers and the degree of masculinity or femininity in the voice were investigated in a perceptual protocol in which listeners heard each speaker reading the first 2 sentences of the Rainbow Passage. Spencer also correlated the perceptual results with measures of SFF. Male and female control subjects, as well as transgendered subjects, were included in the listening protocols. Results indicated that 4 of the 8 transgendered subjects were perceived as female with 70% or better accuracy by listener subjects. However, of those 4 subjects, 3 received low femininity ratings, lower than any of the female control subjects and the fourth received a rating only slightly above the lowest-rated female control subject. Average SFF of the 4 transgendered subjects perceived as female was not given by the investigator, although she noted that all of these individuals had a SFF above 160 Hz. The range of SFFs in the female-perceived group appeared to be 165-209 Hz.

Other recent research has focused on pitch and related intonation patterns in changing the vocal gender identification of male-to-female transgenders. Wolfe, Ratusnik, Smith, and Northrop[4] recorded responses to questions about home and work from 20 male-to-female transsexuals in various stages of the reassignment process, including some who were still living as men. A randomized tape was then prepared, which included the samples of the transgendered subjects plus 10 biological males and 10 biological females. Ten speech students judged the gender of each speaker. A second group of listeners ($N = 8$) rated each speech sample on a 7-point feminine-masculine scale, with 1 representing an extremely feminine voice and 7 representing an extremely masculine

voice. Results of the Wolfe et al study revealed a mean SFF of 172 Hz for the group of transgendered individuals perceived as female, with SFFs ranging from 156 Hz to 195 Hz. The mean SFF for the transgendered individuals perceived as male was 118 Hz, with SFFs ranging from 97 Hz to 140 Hz. Further, Wolfe et al found that the female-perceived group had downward pitch inflections that were significantly less low in frequency than those of the male-perceived group, but had a higher percentage of both upward and downward inflections, indicating more vocal variability in the female-perceived group. Although Wolfe et al measured the extent of upward and downward inflections in terms of semitones, they did not give Hz values for upper and lower limits of SFF. Thus, a SFF range in Hz cannot be specified from this study and the importance of range characteristics in gender identification cannot be determined.

The results of Wolfe et al[4] study contrast sharply with those reported by Andrews and Schmidt[5] in a study of male versus female voice production in a group of male cross-dressers. Cross-dressers are different from transgendered individuals in that cross-dressers usually do not want to permanently change their gender; however, male cross-dressers do want to be recognized as female at least some of the time. Andrews and Schmidt examined the perceptual and acoustic characteristics of 11 males who described themselves as cross-dressers. Each subject was recorded twice; once in a feminine mode and once in a masculine mode. Eighty-eight listeners rated each of the 22 speech samples (11 subjects $\times$ 2 presentation modes) on 18 voice quality perceptual scales. Results revealed that listeners heard significant differences between feminine and masculine speech presentations across 11 speakers and 18 perceptual scales. The most significant differences between presentation modes were on the feminine-masculine perceptual scale and the high-low perceptual scale. In "male mode," speakers were perceived as more masculine; they were also perceived to have a lower pitch. In "female mode," speakers were perceived as more feminine, and to have a higher pitch. This finding is somewhat consistent with Wolfe et al, who found that female-perceived transgenders had higher SFFs. However, in the Andrews and Schmidt study, acoustic data did not reveal a marked difference between SFFs in the two modes. "Male mode" SFFs in the Andrews and Schmidt study averaged 119 Hz, whereas "female mode" SFFs averaged 135 Hz, compared to Wolfe et al's male-perceived transgenders at 118 Hz and female-perceived transgenders at 172 Hz.

One possible reason for the discrepancy between studies is that Wolfe et al[4] looked at listener perceptions of the speakers as "male" or "female," as well as ratings of "masculine" versus "feminine." Andrews and Schmidt[5] looked only at rating scale judgments of femininity-masculinity. It is possible that listeners perceived *all* of Andrews and Schmidt's subjects as males, with the distinction that some speakers sounded like more masculine males and some sounded like more feminine males.

The research cited above suggests that SFF is particularly important to listeners, both in discriminating male from female speakers and in judging "masculinity" and "femininity" on a rating scale. In general, voices with higher SFFs are rated to be more "feminine." However, to actually change gender perception from male to female, a SFF cutoff point in the 156-160 Hz region appears to be crucial. Both Spencer[3] and Wolfe et al[4] found that speakers with SFF's above this point were perceived as female, while those below were perceived as males.

Contributions from formant frequencies and frequency variability may also be important in changing gender perception. For example, Mount and Salmon[2] found that in spite of a SFF of 210 Hz, their client was not perceived as a female until her vowel formants in the F_2 region began to increase in frequency as well. Because Mount and Salmon used only one subject and no formal perceptual protocol, however, their results must be considered preliminary. From the Wolfe et al study, the significant differences found in extent of downward intonations, percentage of upward intonations, and percentage of downward shifts in the male-perceived group compared to the female-perceived group lend support to the idea that frequency variability is also a cue for gender perception. However, typical Hz values for highest and lowest frequencies for male- and female-perceived transgenders are still not known.

The purpose of this study was to further investigate the importance of SFF, upper limit of SFF, lower limit of SFF, SFF range, intonation patterns, and vowel formant frequencies in gender identification by com-

paring the voices of male-perceived and female-perceived male-to-female transsexuals on those acoustic parameters. In addition, correlations between ratings of femininity-masculinity and the acoustic parameters were planned. Of particular interest was further exploration of the SFF "dividing line" of 156-160 Hz between male-perceived versus female-perceived transgenders found by Spencer[3] and Wolfe et al,[4] the intonational differences found by Wolfe et al, and potential formant frequency differences between male- and female-perceived transgenders suggested by the results of Mount and Salmon.[2] The eventual goal of this type of research is to provide data regarding which aspects of a transsexual's voice and speech must be modified to achieve identification as the reassigned gender.

METHOD

Subjects

Fifteen individuals who identified themselves as male-to-female transsexuals served as speaker subjects. For the purpose of this study, a "transsexual" or "transgendered" person was defined as an individual who was in the gender transition process and under professional supervision. Some participants had completed their sex reassignment surgery; some had not yet had their surgery but were living as women full-time; and others were still living as men. Subjects were recruited from the voice clients of the University of Wisconsin-Milwaukee Speech and Language Clinic, personal contacts, and from a local transsexual support group. The transsexual subjects' ages ranged from 20 to 63 years, with a mean of 45 years, 2 months. Their mean height was 5 feet, 10 inches.

Nine control subjects, 3 biological males and 6 biological females, were matched to the transsexual subjects by age and height. Control subjects were required to be within 2 years in age and 2 inches in height of one of the transsexual subjects. Similarity between transsexual and control subjects was considered important because of well-documented age-related changes that occur in fundamental frequency,[6] and also because of the possibility that differences in physical stature might reflect differences in vocal tract size, and hence differences in vocal pitch and vowel formant frequencies. The female control subjects had a mean age of 40 years, 8 months, and a mean height of 5 feet, 9 inches. The male control subjects had a mean age of 50 years, 4 months, and a mean height of 5 feet, 10½ inches. Only 3 male control subjects were used due to the fact that 9 of the transgendered subjects were judged by the investigators to use a voice consistent with their biological gender (male). Therefore, to keep the array of voices balanced (in the judgment of the investigators) between those sounding "male" versus those with a more "female" sound, fewer male control subjects were used.

Speech and voice samples

Subjects were individually seated in a sound-treated Industrial Acoustic Company (IAC) booth and instructed to use a comfortable conversational intensity level for each of the 3 speech tasks (reading and 2 vowel prolongations). Transsexual subjects were given the additional instruction to use their best feminine voice.

The first task consisted of reading the Rainbow Passage.[7] Prior to recording, all subjects were asked to read the passage several times. After they demonstrated the ability to read it fluently and naturally, mouth-to-microphone (Radio Shack 33-3007) distance was set and maintained at 1 inch, through the use of a positioning device against which subjects rested their foreheads. The first 3½ to 4 sentences were then digitally recorded at 22 kHz within a 15-second time frame, using the recording subroutine of Dr. Speech (version 3.0, Tiger Electronics, Inc.) for later acoustic analysis. Simultaneously, the subjects' samples were also recorded on a Marantz PMD 221 audiocassette recorder for later perceptual analysis.

Next, subjects were instructed in the second and third tasks, to prolong the vowels /i/ and /ɑ/ each for 5 seconds, and given opportunities for practice. Participants were signaled to stop after each 5-second recording was collected. The same digitization rate and time frame were employed. As with the sentence samples, both digital and audio recordings of the vowels were made simultaneously.

Perceptual protocols

An experimental tape was prepared, which included the prolonged vowels /i/ and /ɑ/ followed by the second and third sentences of the Rainbow Passage

for each of the 15 transsexual subjects and 9 control subjects. To permit reliability assessment, each speaker's entire sample (2 vowels plus 2 sentences of the Rainbow Passage) was included twice on the experimental tape, for a total of 48 samples. The order of speakers on the experimental tape was determined by a random numbers table; however, consecutive recordings of the same speaker were avoided. Each sample was preceded by a unique identification number and followed by a 5-second response time. In addition, a practice tape was also prepared. This tape consisted of the second and third sentences of the Rainbow Passage for each of the 15 transgendered speakers and 9 controls recorded without identification numbers or pauses between speakers.

Listener subjects included 20 normal hearing undergraduate students recruited from various psychology classes, aged 18-34 years, with a mean age of 21 years. Prior to their participation in this study, listeners were required to pass a pure-tone screening at 20 dB SPL, administered in an IAC booth.

After passing the hearing screening and providing informed consent, listener subjects were first instructed to simply listen to the practice tape to familiarize themselves with all of the voices. Next, participants listened to the experimental tape and completed a rating procedure. All stimuli were presented in an IAC booth through headphones. Tapes were played on a Marantz PMD 221 audiocassette recorder set at a comfortable listening level. Listeners were instructed to identify each speaker as a male or a female, estimate the speaker's age, and then rate the femininity-masculinity of each voice on a 7-point rating scale. On each 7-point scale, 1 represented a very feminine voice for the selected gender of the speaker and 7 represented a very masculine voice. Thus, a perceived male voice could be rated either as very masculine for a male or very feminine for a male; the same could be done for a perceived female voice. The emphasis of the listening task was placed on the feminine/masculine scale and age estimation in the hope of disguising gender identification as the primary decision. All listeners were unaware of the purpose of the study or the use of transgendered speakers. Each listener subject was paid for his or her participation at the conclusion of the perceptual protocol.

Acoustic analysis

Acoustic analyses were done to compare the voices perceived as male to voices perceived as female. Each speaker was listened to and judged twice by each listener, and only transgendered subjects that were perceived as male or female at least 70% of the time were retained for further analysis. The selected subjects' readings of the Rainbow Passage were analyzed for speaking fundamental frequency (SFF), SFF range, and intonation patterns. The prolonged vowel samples were analyzed for the first 3 vowel formants. Specific analysis procedures were as follows.

SFF measures

Mean, upper limit, and lowest lower limit of SFF (in Hz) and range (in semitones, or ST) were calculated for the second and third sentences of the Rainbow Passage using the Speech Analysis subroutine of Dr. Speech. This was accomplished first by using the cursor to block the target portion of the subject's recorded Rainbow Passage, and then selecting *Pitch Extraction* from a program-provided menu to be performed on the blocked sample. The Speech Analysis subroutine permits the user to specify a range of frequencies to be included in the pitch extraction procedure: for the present study, a lower limit of 65 Hz (the default) and an upper limit of 600 Hz were used. The upper limit of 600 Hz was selected because, in the experience of the investigators, the highest frequencies utilized by speakers in this study were unlikely to go above this point. After completion of the pitch extraction procedure, *Statistical Analysis* was selected from the program's menu, which provided the maximum, minimum, and average frequency present in the blocked sample. These values were recorded on a data sheet by the investigators.

Intonation measures

Intonation was evaluated for the Rainbow Passage sample using the following parameters: extent of intonation shift (upward and downward) in ST and number of upward and downward intonation shifts. An intonation shift was defined as a change in frequency, with or without interruption of phonation, of at least 2 semitones. Analysis of intonation shifts was accomplished by first visually inspecting the pitch trace on a frequency-by-time display output by the pitch extraction procedure previously performed to

determine SFF. Intonation shifts were identified by placing a cursor at the beginning of the sample, noting the frequency, and then placing the cursor at the top (or bottom) of the rising (or falling) pitch trajectory that proceeded from the initial point, and again noting the frequency. If the 2 frequencies were 2 or more semitones apart, a frequency shift in the proper direction (up or down) was counted, and its extent in semitones recorded. If the 2 frequencies were not 2 semitones apart or more, no data were recorded. The highest (or lowest) point of the next frequency contour was identified, and the procedure was repeated until all of the frequency shifts in the sample had been analyzed.

Formant frequency measures

The vowels /ɑ/ and /i/ from the vowel prolongation task were analyzed. The first three formants (F_1, F_2, and F_3) of each vowel were computed using the speech analysis subroutine of Dr. Speech. For this analysis, a Hamming window was selected, with a pre-emphasis (designed to boost the less intense, higher frequencies) of 90 (the default). As with the Rainbow Passage, the target portion of each vowel was blocked using the program's cursor. In this case, the target portion consisted of the most stable middle 3-second section of the vowel, excluding onset and offset. The specific procedure selected from the analysis menu was *Long-Term Power Spectra with LPC (Linear Predictive Coding) Analysis.*

This option was selected because it provides both the spectrum of the vowel, showing all the harmonics of the fundamental frequency, as well as a mathematical estimation of the peaks present in that spectrum. The frequencies corresponding to the peaks, or formant frequencies, were read off the display provided by the program. In the event that the program did not identify formants within the expected frequency ranges, formant frequencies were manually estimated from the program's spectral analysis display by placing the cursor at the highest amplitude harmonic in the expected range and recording the resulting frequency. Two researchers independently analyzed any sample for which formants were not provided by the program and compared their results. Exact matches in formant frequency estimations were recorded. Minor discrepancies in frequencies between the 2 investigators were averaged and recorded. No major discrepancies occurred.

Statistics

The procedures described above resulted in a total of 14 acoustic measures for each selected subject: mean SFF, SFF range, upper limit of SFF, lower limit of SFF, mean upward shift in ST, mean downward shift in ST, number of upward shifts, number of downward shifts, F_1 of /i/, F_2 of /i/, F_3 of /i/, F_1 of /ɑ/, F_2 of /ɑ/, and F_3 of /ɑ/. There were also 2 perceptual judgments for each selected subject: gender, and femininity-masculinity rating.

Significant differences between male-to-female transsexuals vocally perceived as female and those vocally perceived as male were calculated using Mann Whitney *U* tests for all acoustical measures, as well as median femininity-masculinity rating. Differences at $P \leq .05$ were considered significant. In addition, Spearman rank-order correlation coefficients were calculated for the feminine-masculine rating and each of the remaining 14 dependent variables. Correlations were examined for the combined perceived male-perceived female groups. Correlations at $P \leq .05$ were considered significant.

RESULTS

Of the 15 transsexual speaker subjects included in the perceptual protocol, 10 were identified as male speakers, 3 were identified as female speakers, and 2 were not consistently perceived as male or female at least 70% of the time. Of the male-rated speakers, 6 were identified as male 100% of the time, 3 were identified as male 97.5% of the time, and 1 speaker was identified as male 90% of the time. For speakers rated as female, 2 were identified as female 92.5% of the time and 1 was identified as female 90% of the time. Two groups were formed for further acoustical and perceptual analysis: subjects perceived as male (N = 10) and subjects perceived as female (N = 3). Of the speakers whose gender was not judged consistently, one was identified as female 37.5% of the time; the other was identified as female 57.5% of the time. These subjects were not used for further statistical comparisons or correlations.

Listener reliability

Each voice was presented twice on the experimental tape; thus each listener rated each speaker as male

or female twice. For the 13 transgendered subjects of primary interest, this resulted in a total of 520 judgments (26 gender identifications × 20 listeners), or 260 pairs of judgments. At the same time gender was determined, listeners were also asked to rate the femininity or masculinity of the voice for the selected gender. Thus, 260 pairs of femininity-masculinity ratings were also available for reliability analysis.

In terms of gender identification, listener subjects rarely changed their judgment of a speaker from male to female, or vice versa. Of 260 pairs of judgments, there were only 13 reversals, representing only 5% of the total pairs judged. All control subjects were identified as the correct gender 100% of the time, except for one female control subject who was identified as female 82% of the time. Interestingly, this subject had the lowest speaking fundamental frequency (SFF), 165 Hz, of the female control group.

In addition to looking at the intrajudge reliability of gender perception, intrajudge reliability of feminine-masculine ratings was examined. Because the feminine-masculine ratings were dependent on the selected gender, the 13 incidents of a listener perceiving a speaker as male on one presentation and female on another were excluded. This resulted in 247 pairs of ratings. A Spearman rank-order correlation coefficient between the feminine-masculine rating of each listener's first and second rating of each speaker was calculated to be $r_s = 0.76$ ($P < 0.001$). This indicates a highly significant, moderate to high level of intrajudge reliability for the feminine-masculine ratings.

Comparisons between groups

A visual inspection of the data presented in Tables 1, 2, and 3 reveals differences between the perceived-female and the perceived-male groups for many of the acoustic measures. Speaking fundamental frequency (SFF), upper and lower limits of SFF, and all vowel formants were consistently higher for the female-perceived group compared to the male-perceived group. In addition, female-perceived subjects had a greater range (more semitones) in their downward pitch shifts, and a greater number of upward shifts in pitch.

The Mann Whitney *U* test was used to identify significant differences between the 2 speaker groups on each of 15 dependent variables. The Mann Whitney *U* test, a nonparametric procedure, was used because several of the assumptions required for parametric statistics could not be met. First, the size of the female-rated group (N = 3) was not sufficiently large. Second, because of the small size of that group, the variances of the 2 transgendered groups were not similar enough to warrant the use of parametric statistics. Further, to correlate median femininity-masculinity rating (an ordinal variable) with other acoustic variables, a nonparametric procedure was required.

As shown in Table 1, significant differences ($P < 0.05$) were found on speaking fundamental frequency (SFF) and upper limit of SFF. None of the other comparisons achieved statistical significance. As expected, speakers perceived as female had significantly higher SFFs than speakers perceived as male. The

TABLE 1. *Means, Ranges, and Mann Whitney U Results for Transgendered Subjects Perceived as Female Compared to Transgendered Subjects Perceived as Male on Parameters Relating to Speaking Fundamental Frequency (SFF). All Measures are Given in Hz Except Range, Which is Given in Semitones (ST).*

	Perceived Female (N = 3)		Perceived Male (N = 10)		
Measure	Mean	Range	Mean	Range	Significance
SFF (Hz)	187	164-199	142	112-181	$P = 0.03$
SFF upper limit (Hz)	301	276-320	222	150-279	$P = 0.02$
SFF lower limit (Hz)	138	115-168	105	79-138	$P = 0.06$
SFF range (ST)	13.7	11-17	12.9	9-19	$P = 0.67$

TABLE 2. *Means, ranges, and Mann Whitney U results for transgendered subjects perceived as female compared to transgendered subjects perceived as male on vowel formant frequencies measured in isolated vowels in Hz.*

	Perceived Female (N = 3)		Perceived Male (N = 10)		
Measure	Mean	Range	Mean	Range	Significance
F_1 of /ɑ/	811	732-926	781	646-870	$P = 0.80$
F_2 of /ɑ/	1341	1205-1507	1208	1033-1311	$P = 0.18$
F_3 of /ɑ/	2793	2670-3040	2662	2283-3470	$P = 0.39$
F_1 of /i/	273	215-345	235	173-322	$P = 0.43$
F_2 of /i/	2382	2326-2412	2313	1981-2541	$P = 0.39$
F_3 of /i/	3065	2821-3230	2940	2627-3475	$P = 0.40$

TABLE 3. *Means, ranges, and Mann Whitney U results for transgendered subjects perceived as female compared to transgendered subjects perceived as male on intonation measures and female-male ratings. Mean upward and downward shifts are reported in semitones (ST).*

	Perceived Female (N = 3)		Perceived Male (N = 10)		
Measure	Mean	Range	Mean	Range	Significance
Mean upward shifts (ST)	4.9	3.4-6.1	4.7	3.2-5.8	$P = 0.40$
Mean downward shifts (ST)	5.9	4.8-7.2	4.9	3.5-6.0	$P = 0.15$
Number of upward shifts	18.0	14-23	17.1	13-21	$P = 0.86$
Number of downward shifts	18.3	15-23	18.2	13-21	$P = 1.00$
Feminine-masculine rating	4.0*	2-4	4.5*	2-5	$P = 0.29$

* Median values instead of means.

mean for the female-perceived group was 187 Hz, with individual speakers' SFFs ranging from 164 Hz to 199 Hz. The results of this group were similar to the female control speakers, who also had a mean SFF of 187 Hz, with individual speakers' SFFs ranging from 165 Hz to 221 Hz. Speakers perceived as male had a group mean of 142 Hz, with a range of SFFs from 112 Hz to 181 Hz. Thus, there was an unexpected overlap in SFF between the male-rated and female-rated subjects extending from 164 Hz to 181 Hz.

Speakers perceived as female also had significantly higher SFF upper limits than those perceived as male. Speakers perceived as female had a mean SFF upper limit of 301 Hz, with individual upper limits encompassing a range of 276-320 Hz. It is interesting to note that the female control subjects had a mean SFF upper limit of only 258 Hz, with a range of 216 Hz to 320 Hz. In contrast, transgendered speakers perceived as male had a mean SFF upper limit of 222 Hz, with a range of 150 Hz to 279 Hz (see Table 1). Again, some overlap was seen between the male- and female-perceived subjects. No other acoustic measures showed significant differences between groups.

A significant difference was not found between female- and male-identified subjects on the femininity-masculinity ratings. Female-identified subjects had a median rating of 4, and male-identified subjects had

a median rating of 4.5 (on a scale where 1 anchored the feminine end and 7 anchored the masculine end of the continuum). Thus, female-rated subjects achieved only a slightly higher feminine rating than the male-rated subjects.

Correlations

As shown in Table 4, Spearman rank-order correlation coefficients between the feminine-masculine ratings and each of the remaining 14 acoustic measures were computed for the 13 transgendered subjects included in this study. Correlations among 13 pairs are considered significant at the 0.05 level if $r^s = 0.48$.[8] The upper limit of SFF was the only parameter to achieve significance, with a negative correlation of $r^s = -0.61$. This meant that transgendered subjects with higher frequency upper SFF limits were in general perceived as more feminine (a rating closer to 1) than transgendered subjects with lower frequency upper SFF limits, who were perceived as more masculine.

Two other parameters that obtained moderate but not statistically significant correlations with femininity-masculinity ratings were range in semitones, or ST ($r_s = -0.43$) and F_2 of /i/ ($r^s = -0.42$). As range in ST and as the F_2 of /i/ increased, listeners tended to rate the speaker as more feminine. It should be noted, however, that the speaker was not necessarily more likely to be identified as female, just more likely to be rated feminine, regardless of gender.

TABLE 4. *Spearman Rank-order Correlation Coefficients Between Femininity(1) and Masculinity(7) Ratings and Each of the Acoustic Parameters for All Transgendered Subjects (N = 13).*

Measure	Correlation With Femininity-Masculinity Rating
SFF (Hz)	−0.36
SFF upper limit (Hz)	−0.61*
SFF lower limit (Hz)	−.020
SFF range (ST)	−0.43
F_1 of /ɑ/	−0.36
F_2 of /ɑ/	−0.19
F_3 of /ɑ/	−0.38
F_1 of /i/	0.01
F_2 of /i/	−0.42
F_3 of /i/	−0.16
Mean upward shift (ST)	−0.13
Mean downward shift (ST)	−0.22
Number of upward shifts	−0.23
Number of downward shifts	−0.29

*Significant at the .05 level or better.

DISCUSSION

The purpose of the present study was to explore significant differences between transgendered speakers identified as male and those identified as female for speaking fundamental frequency (SFF), upper and lower limits of SFF, SFF range, vowel formants, and intonation, as well as to identify any significant correlations between feminine-masculine ratings and those acoustic measures. Using the Mann-Whitney U test, significant differences ($P \leq 0.05$) were found for SFF and upper limit of SFF. Subjects perceived as female had a higher SFF and higher upper limit of SFF than subjects perceived as male. Differences on other parameters did not achieve statistical significance; however, female-perceived subjects had consistently higher SFF lower limits, higher vowel formant frequencies for isolated productions of /i/ and /ɑ/, a greater number of upward intonation shifts, and a greater range (in ST) of downward intonation shifts.

Spearman rank-order correlations revealed that the upper limit of SFF and median scores of femininity-masculinity were significantly related. Again, the higher in frequency the upper limit of SFF, the more likely the speaker was to be given a feminine rating. Range in ST and the second formant frequency of the vowel /i/ also moderately correlated with femininity-masculinity scores, but not to a significant degree.

A risk in comparing the data of two groups selected through the perceptual judgments of untrained listeners is that there is no guarantee of equally distributed groups. A greater number of transsexuals perceived as female would have produced more representative data, provided more valid comparisons with the larger male-perceived group, and permitted the use of parametric statistics. All of these factors

limit the generalizations, that can be drawn from this study. In addition, the transsexual subjects varied greatly in their progress with gender reassignment. Some subjects had not yet begun to live as females full-time, while others had been living as female for more than a decade. Also, some subjects had never had professional speech or language intervention, whereas others had an extensive history of speech and language therapy. Finally, another variation among the transgendered subjects was age, ranging from 22 years to 63 years. It is possible that listeners use different criteria in determining the gender of speakers they perceive to be young versus speakers they perceive to be old. Better control of these factors (gender reassignment status, speech therapy history, and age) might permit a clearer picture of the variables that differentiate transgendered subjects perceived as female from those perceived as male.

Despite the limitations of the present study, there were several areas of agreement between this study and previous research. Wolfe et al[4] reported a mean SFF for transgendered subjects perceived as female of 172 Hz, with a range in individual SFFs of 156 Hz to 195 Hz; Spencer[3] found a range of individual SFFs of approximately 165 Hz to 209 Hz. The present results of a mean of 187 Hz for the female-perceived group, with a range in individual SFFs of 164 Hz to 199 Hz, were clearly similar to other studies. In addition, no female-perceived subject in the present study had a SFF below the 156-160 Hz dividing line established in previous literature. Thus, the present study supported the conclusion of the Spencer and Wolfe et al studies that a SFF above 156-160 Hz is necessary for a transgendered individual to be perceived as female.

However, the present study did not support the finding of Wolfe et al[4] and Spencer[3] that an SFF above 156-160 Hz was *sufficient* to be perceived as a female. In both previous studies, all transgendered individuals with SFFs above a 156-160 Hz dividing line were perceived as female and all subjects with SFFs below that point were perceived as males. This study did not replicate these findings. Although there were significant differences between groups in mean SFF, a few male-perceived subjects had SFFs well in excess of 160 Hz.

There were also differences between the results of the present research and those of Wolfe et al[4] in terms of intonational contrasts between male-perceived and female-perceived transgenders. Wolfe et al found significant differences between groups on 5 intonation parameters: extent of downward intonations, percentage of upward and level intonations, and percentage of downward and level shifts. This study found no significant differences in intonational analysis between subjects perceived as male and those perceived as female. One possible reason for this discrepancy is the difference in intonational analysis. Wolfe et al differentiated between *intonations*, or pitch change during connected vocalization and *shifts*, pitch changes that occurred between the end of one vocalization and the beginning of another. Due to the difficulty in identifying interword versus intraword pitch variations, such a distinction was not made in this study. In addition, Wolfe et al used spontaneous speech samples, while the present study used the second and third sentences of the Rainbow Passage. It is possible that when a subject is more "engaged" in the speech he or she is formulating, a greater variety of intonational patterns are produced and these patterns may be more salient in differentiating male and female speakers.

A number of surprising findings emerged from this study. First, among the initial group of 15 transsexual subjects, it was expected that 6 of the subjects, with SFFs of 170 Hz and above, would be perceived by listeners as female. The 170 Hz figure was taken from Wolfe et al,[4] whose female-perceived group had a mean SFF of 172 Hz. Instead, in the present study, only 2 of those 6 speakers (plus 1 with a lower SFF) were identified as female. Two were not consistently identified as either gender, and 2 were, in fact, identified as males 90% or more of the time.

In light of the reliability with which listeners judged the speakers to be male or female, it was also unexpected that so few differences in acoustic measures would emerge between the perceived-male and perceived-female groups. In particular, the results of Mount and Salmon[2] suggested that formant frequencies would be a salient cue in gender differentiation. Mount and Salmon concluded that their subject was not perceived as female, despite attaining a SFF of 210 Hz, until she was able to raise her second formant frequencies. Yet, in the present study, none of the comparisons between vowel formants for the perceived-male and perceived-female group attained significance.

It was also surprising to see so little difference between the median femininity-masculinity rating of the female-perceived subjects compared to the male-perceived subjects. Based on Bralley et al[1] and Andrews and Schmidt,[5] it was expected that transgendered subjects perceived as female would also be perceived as more feminine than transgendered subjects perceived as male. In fact, very little difference in median rating was seen between the two groups. It appears that it is not possible to generalize a listener's judgment of a speaker's femininity or masculinity to the same listener's gender identification of the speaker.

Why did these unexpected results occur? It is possible that the listeners had a predisposition to identify voices reading neutral, nonemotional sentences as male. That is, if a voice presented with some characteristics of a female voice and some characteristics of a male voice, listeners may have had a tendency to attend to the male characteristics and identify the voice as belonging to a male (although perhaps a feminine-sounding male). This was noted perceptually by the senior author, who found that, for some of the experimental voices, the impression was of a female speaker, except for a single word or perhaps phrase that had a "male" sound to it. These experimental voices were consistently identified by listeners as belonging to a male speaker. Perhaps this study failed to find significant differences between the perceived-male and the perceived-female groups because the investigators were not measuring the specific segment of each speaker's sample on which listeners based their gender identification judgment.

In addition, a variety of variables that affected listeners' judgments of gender may not have been examined. For example, in terms of physiological differences in voice production, Linville[6] found that young men displayed fairly low incidence of glottal gap, whereas young women displayed a significantly higher incidence of glottal gap during stroboscopic assessment. Sodersten, Hertegard, and Hammarberg[9] found a posterior chink glottal configuration in 61% of their female subjects. These physiological differences in voice production may result in acoustic attributes such as breathiness that are associated with masculinity or femininity in a voice and may have affected listener judgments in the present study. Further, Sorensen and Horii[10] have suggested that differences in jitter and shimmer values exist between male and female voices. Because these parameters were not examined in the present study, it is not known how they affected the final results.

Which vocal characteristics are most important in gender identification and how successfully can the perceived gender of a speaker be changed? Based on present and past research results, the answer to the first question appears to be SFF, upper limit of SFF, lower limit of SFF, intonational variability, and possibly resonance characteristics, although the latter is less well-researched at this point. Because the present study examined only listener judgments of read sentences, it is not possible to determine the effects of word choice, perceived affect, intensity, or durational characteristics. In spontaneous speech samples, these aspects of speech most likely would be more variable among subjects and might reveal differences between the perceived-male and perceived-female groups. As for the success of changing a speaker's gender identification through vocal cues alone, it is clear that success for some individuals is possible, even when the speech sample is neutral and there are no visual or contextual cues. However, male vocal characteristics appear to be more salient than female vocal characteristics; thus a thorough and multi-dimensional voice assessment and treatment plan is needed for a transgendered client attempting to acquire a female-perceived voice. Finally, further research is needed to determine how gender is identified when additional vocal and semantic cues are available and how greater success in altering gender perception for the male-to-female transsexual population may be attained.

REFERENCES

1. Bralley RC, Bull GL, Gore CH, Edgerton MT. Evaluation of vocal pitch in male transsexuals. *J Comm Disorder.* 1978;11:443-449.
2. Mount KH, Salmon SJ. Changing the vocal characteristics of a postoperative transsexual patient: a longitudinal study. *J Comm Disorder.* 1988;21:229-238.
3. Spencer LE. Speech characteristics of male-to-female transsexuals: a perceptual and acoustic study. *Folia Phoniatr.* 1988;40:31-42.
4. Wolfe VI, Ratusnik DL, Smith FH, Northrop GE. Intonation and fundamental frequency in male-to-female transsexuals. *J Speech Hear Disorder.* 1990;55:43-50.
5. Andrews ML, Schmidt CS. Gender presentation: perceptual and acoustical analyses of voice. *J Voice.* 1997;11:307-313.

6. Linville SE. The sound of senescense. *J Voice.* 1996;10: 190-200.
7. Fairbanks GE. *Voice and Articulation Drillbook.* 2nd ed. New York, NY: Harper; 1960.
8. McCall RB. *Fundamental Statistics for Behavioral Sciences.* 6th ed. Fort Worth, Tex: Harcourt Brace; 1994.
9. Sodersten M, Hertegard S, Hammarberg B. Glottal closure, transglottal airflow, and voice quality in healthy middle-aged women. *J Voice.* 1995;9:182-197.
10. Sorenson D, Horii Y. Frequency and amplitude perturbation in the voice of female speakers. *J Comm Disorder.* 1983;16:57-61.